MATHEMATICS OF COMPUTATION, VOLUME 29, NUMBER 132
OCTOBER 1975, PAGES 1046–1050

Polynomial Evaluation Schemes

By A. C. R. Newbery*

Abstract. An attempt is made to define a polynomial evaluation algorithm that is more resistant to accumulated round-off error than the schemes of Horner and Clenshaw under conditions of floating-point arithmetic. An algorithm is presented which generally compares favorably with both. Some suggestions are made, which could plausibly lead to substantial further improvements.

In [1] it was established that the total effect of rounding error for Horner's (nested multiplication) evaluation scheme is bounded by $(\epsilon + n\sigma)\,\tilde{P}(|\alpha|)/(1 - n\sigma)$, where the arithmetic is performed in floating-point, ϵ is the smallest number which significantly adds to 1, $\sigma = \epsilon(2 + \epsilon)$, and $\tilde{P}(x) \equiv \Sigma_0^n |p_r| x^r$. It was noted that both the theoretical and the empirically observed errors vary steeply with the magnitude of the argument. In view of this, it seems appropriate to look into the possibility of replacing a given polynomial evaluation problem by an equivalent one for which the argument has a smaller magnitude. We are presupposing a situation in which the overhead costs (for transforming one problem into another) may be neglected, e.g. where the polynomial approximates an elementary function on a library tape.

Let the given problem be to evaluate $P(x) \equiv \Sigma_0^n p_r x^r$ at any argument in $[-1, 1]$. It was observed in [1] that any finite-domain problem can be scaled into this form. For convenience we assume $n = 2n'$. (The analysis for odd-degree polynomials is only trivially different.) First, we split P into its even and odd components $P(x) \equiv E(x^2) + x\Phi(x^2)$, where E and Φ are polynomials of degree n', $n' - 1$ in the argument x^2; secondly, by writing $x^2 = \frac{1}{2} + t$ with $|t| \leqslant \frac{1}{2}$, we can rewrite $E(x^2)$ and $\Phi(x^2)$ as $\hat{E}(t)$ and $\hat{\Phi}(t)$ with $|t| \leqslant \frac{1}{2}$. At this point, it is clear that we can reduce the maximal argument size from 1 to $\frac{1}{2}$, which is generally advantageous; but we need to ensure that this advantage will not be outweighed by a growth in coefficient size which might occur when we transform the polynomials E, Φ into $\hat{E}$, $\hat{\Phi}$. For simplicity we shall only look at the transformation from E to $\hat{E}$ and we shall take the global error bound for $P(x)$ as $2n\epsilon\tilde{P}(1)$. (We have neglected some small quantities and assigned to $|\alpha|$ its "worst" possible value of 1.)

Defining $E(s) \equiv \Sigma_0^{n'} e_r s^r \equiv \Sigma_0^{n'} p_{2r} s^r$ and $\hat{E}(t) \equiv \Sigma_0^{n'} \hat{e}_r t^r$, the following relation holds between the vector $\hat{e} \equiv (\hat{e}_0, \hat{e}_1, \ldots, \hat{e}_{n'})^T$ and $\bar{e} \equiv (e_0, e_1, \ldots, e_{n'})^T$:

(1) $\hat{e} = M\bar{e}$, where

Received April 4, 1974.
AMS (MOS) subject classifications (1970). Primary 65G05, 68A10.
Key words and phrases. Error analysis, polynomials.
*The author gratefully acknowledges support for this work by the Science Research Council of Great Britain.

$$M = \begin{bmatrix} 1 & 1/2 & 1/2^2 & 1/2^3 & 1/2^4 & \cdots \\ & 1 & 1 & 3/4 & 1/2 & \cdots \\ & & 1 & \frac{C_{3,2}}{2} & \frac{C_{4,2}}{2} & \cdots \\ & & & 1 & \frac{C_{4,3}}{2} & \cdots \\ & & & & 1 & \cdot \\ & & & & & \cdot \\ & & & & & \cdot \end{bmatrix}.$$

If we adopt the convention that the upper-triangular matrix M has a "zeroth" row and column, then its entries are given by $m_{ij} = C_{j,i}/2^{j-i}$, where $C_{j,i}$ is a binomial coefficient

The simplified error bound for evaluating $\hat{E}(t)$ in the worst case where $|t| = \frac{1}{2}$ is $2\epsilon n' \Sigma_0^{n'} |\hat{e}_r| (\frac{1}{2})^r = 2\epsilon n' \|G\hat{e}\|_1$ where $G = \operatorname{diag}\{1, \frac{1}{2}, \ldots, \frac{1}{2^{n'}}\}$. We have to compare this with that part of the Horner method error which arises from even-order coefficients namely $2n\epsilon\Sigma_0^n |p_{2r}| = 4n'\epsilon\|\bar{e}\|_1$. Removing common factors, the comparison is between $\|G\hat{e}\|_1$ and $2\|\bar{e}\|_1$. From (1) we obtain

$$\|G\hat{e}\|_1 = \|GM\bar{e}\|_1 \leqslant \|GM\|_1 \|\bar{e}\|_1. \tag{2}$$

It may be observed that GM is column-stochastic, so that $\|GM\|_1 = 1$; and, therefore, $\|G\hat{e}\|_1 \leqslant \|\bar{e}\|_1$. We had to beat $2\|\bar{e}\|_1$ in order to break even, so the proposed method always has a factor of at least 2 in its favor. The factor will be exactly 2 only in the event that $\bar{e}$ is an equisign vector, since that is the situation in which the inequality (2) becomes an equality. Having established that the factor favoring the proposed method is at least 2, we now attempt to find an upper bound for it. Letting F denote the factor, which is the ratio of error bounds, we have $F = 2\|\bar{e}\|_1 / \|G\hat{e}\|_1$. Using the fact that $\bar{e} = M^{-1}\hat{e}$, we have

$$F = 2\|M^{-1}\hat{e}\|_1 / \|G\hat{e}\|_1 = 2\|M^{-1}G^{-1}G\hat{e}\|_1 / \|G\hat{e}\|_1 \leqslant 2\|M^{-1}G^{-1}\|_1. \tag{3}$$

We calculate that

$$M^{-1} = [m'_{ij}] = [(-1)^{i+j} C_{ji}/2^{j-i}], \tag{4}$$

under the convention that we have a 'zeroth' row and column. The absolute sum of the kth column of $M^{-1}G^{-1}$ is A_k, where

$$A_k = 2^k \sum_{r=0}^{k} |m'_{rk}| = \sum_{r=0}^{k} C_{k,r} 2^r = 3^k. \tag{5}$$

It follows that $\|M^{-1}G^{-1}\|_1 = 3^{n'}$ and we conclude from (3) that

$$F \leqslant 2(3^{n'}). \tag{6}$$

The bound (6) will be attained if, and only if $\hat{e}_{n'}$ is the only nonzero coefficient of

$\hat{E}(t)$. The error arising from the odd-order coefficients Φ_r are subject to similar bounds except that when $n = 2n'$, the degree of Φ is $n' - 1$. In general, when the parity of n is unrestricted, the ratio of error bounds $\bar{F}$ will be within the range

$$2 \leqslant F \leqslant 2(3^{\bar{n}}), \quad \text{where } \bar{n} = \text{entier}((n-1)/2). \tag{7}$$

We shall denote the proposed method by E. O. (Even-Odd). If we attempt to compare the three methods, Horner, Clenshaw and E. O., on the basis of error bounds (remembering that the actual errors may not be well correlated with the bounds), we note that the asymptotic error bound ratio favoring Clenshaw over Horner is in the range $[1, (1+\sqrt{2})^{n+1}/2]$ as was shown in [1]. Since $(1+\sqrt{2})^{n+1}/2 > 2(3^{\bar{n}})$, we deduce that Clenshaw is capable of outperforming Horner by a greater margin than E. O. On the other hand, judging by error bounds, E. O. *consistently* outperforms Horner, while Clenshaw does not.

In order to check the validity of the theory, four tests were run. There seems to be no way of defining a set of fair or 'typical' polynomials, so each test was designed to prove a point. While each individual test is admittedly loaded, the set of tests collectively is thought to present a fair picture. The tests were run at the Oxford University Computing Laboratory with 36-bit-mantissa rounded arithmetic. For a given method, polynomial and argument, the "error" is taken to be the absolute difference between single and double precision evaluations *by that method.* Hence our judgement will not be obscured by the fact that, for example, small computational errors might affect the values of $\hat{e}_r$. The argument range $[-1, 1]$ was split into $|\alpha|$ "small" ($<.75$) and $|\alpha|$ "large". In each subrange, 100 equispaced values of α were chosen; and the maximal error for each method was determined. The test polynomials were

P1. $\Sigma_0^{10} x^r/r!$. (This equisign polynomial should theoretically favor Horner against both competitors.)

P2. $T_{10}(x)$, the tenth degree Chebyshev polynomial. (This should favor Clenshaw. It is biassed against Horner because no account is taken of vanishing odd-order coefficients.)

P3. Same as P2, except that the order of the coefficients is reversed. (Biassed against Horner, but otherwise neutral.)

P4. $(1+x)T_{10}(x)$. (Although there are no missing coefficients, this is still biassed against Horner because the signs alternate *in pairs.* The Horner method works best when the signs alternate in ones or are constant.)

The following table exhibits the maximum errors for each method, polynomial and argument range in units of 10^{-12} to two significant figures. The three methods, H, C, E. O. are ranked in order of merit.

Comments on empirical results.

(i) The E. O. method is consistently best for large $|\alpha|$ and consistently second best for small $|\alpha|$. If the results for the two $|\alpha|$ ranges were merged, the E. O. method would be best on all except P3.

(ii) The relatively poor performance of the E. O. method when $|\alpha|$ is small is

explained by the fact that it is at its worst when $|t| = \frac{1}{2}$, and this occurs at $\alpha = 0$. The other two methods do not encounter their worst argument values of ± 1 when $|\alpha| < .75$.

(iii) The theory gives an accurate forecast of when Horner's method will perform well or badly.

(iv) The theory forecasts incorrectly that the Clenshaw method will give best results on P2, but the empirical figures do not stand in very sharp contrast to the theory.

TABLE

	$\|\alpha\| < .75$		$\|\alpha\| \geqslant .75$	
	H	14	E. O.	22
P1	E. O.	17	H	27
	C	25	C	43
	C	33	E. O.	53
P2	E. O.	97	C	150
	H	1300	H	8800
	C	4100	E. O.	1100
P3	E. O.	4600	C	3600
	H	5000	H	13000
	C	30	E. O.	110
P4	E. O.	100	C	280
	H	1800	H	14000

Conclusions. It appears that there is no one method which is better than the others over the whole range of polynomials. At present there is no method significantly better than Horner's when the coefficients are equisign or of strictly alternating sign; however, it has been shown here and in [1] that Horner's method is capable of producing very bad results when these conditions do not hold. In regard to costs, the E. O. method requires $n + 1$ additions and multiplications as compared with n of each for Horner's method; however, in a parallel-arithmetic environment the E. O. method would run about twice as fast. Several questions arise from this study. Firstly, if an advantage can be gained by splitting a polynomial into even and odd parts, why not continue and split the even part and odd part each into two, etc.? Secondly, since we saw in [1] that Clenshaw is a better "defensive" choice than Horner (i.e. Clenshaw at his worst is not as bad as Horner at his worst), would it not be better to evaluate $\hat{E}$ and $\hat{\Phi}$ by Clenshaw's method, rather than Horner's, as was done? Thirdly, since the success of the E. O. method depends on reduction of argument size, would it not be reasonable to generalize it in the following way? The E. O. method replaces $P(x)$ by $\hat{E}(t) = x\Phi(t)$, where $t = x^2 - \frac{1}{2} = (\frac{1}{2})T_2(x)$. Why not try replacing $P(x)$ by $Q(z) + xR(z) + x^2S(z)$, where $z = T_3(x)/2^2$? None of these variants has been tried, and since they are quite numerous it would seem that a systematic study would more appropriately be undertaken by a team rather than by an individual. It seems entirely plausible to believe that within this domain of possible methods, there is one which will consistently

outperform both the established methods while maintaining comparable costs. The E. O. method comes close to doing precisely that, and we have no reason to think it is one of the best methods among those that are envisaged.

Department of Computer Science
University of Kentucky
Lexington, Kentucky 40506

1. A. C. R. NEWBERY, "Error analysis for polynomial evaluation," *Math. Comp.*, v. 28, 1974, pp. 789–793.